CA Journal of **Cultural Analytics**

Understanding Gender and Character Agency in the 19th Century Novel

Matthew Jockers and Gabi Kirilloff

12.01.16

The relationship between character identity and character action is an established topic of literary study. In *Morphology of the Folktale,* Vladimir Propp argues against the separation of "who acts" from "the question of the actions themselves," instead advocating an approach that studies characters according to their functions.[1] Similarly, the notion of the literary archetype, as influenced by both Joseph Campbell and Carl Jung, highlights the way in which what characters *do* influences the way we perceive who characters *are*.[2] The notion of literary stereotypes is also tied to this discussion. As Bamman, O'Connor, and Smith observe in their study of personas in film, stereotypical characters are "defined by a fixed set of actions widely known to be representative of a class."[3] While actions are only one part of a complex network of descriptive tools that authors may use to create characters, [4] they may offer us a useful insight into the way that certain behaviors can align with various character identity traits in literature. A study of character action may serve as a proxy to not only demarcate character types, but also to investigate what behaviors, and *types* of behaviors, were conventionally aligned with different groups of characters. Bamman, Underwood, and Smith (2014) attempt something similar in "A Bayesian Mixed Effects Model of Literary Character." In that work, the authors focus on modeling specific character "personas" by studying semantically related words that occur in proximity to character mentions. The authors note that "articulating what a true 'persona' might be for characters is inherently problematic" and they acknowledge that the "personas learned so far [by their model] do not align neatly with character types known to literary historians."[5] Nevertheless, their study revealed compelling associations between certain personas and certain genres and, more importantly for our current research, that certain personas were "clearly gendered." [6] In noting the later, the authors write that "analysis of latent character types might cast new light on the history of gender in fiction."[7] Our work attempts a more direct and specific study of character agency in the context of character gender. To do this, we explore trends in behavior associated with male and female characters in 3,329 19th century novels.[8]

---

[1] Vladimir Propp, Morphology of the Folktale: Second Edition (Austin: University of Texas Press, 1968), 87.

[2] E.g. Joseph Campbell, The Hero with a Thousand Faces (Princeton, N.J.: Princeton University Press, 1972); C. G Jung and R. F. C. Hull, The Archetypes and the Collective Unconscious (Princeton, N.J.: Princeton University Press, 1980).

[3] Bamman et. al. analyze character personas in a large corpus of plot summaries using metadata obtained from a database (Freebase) of film characters and the actors who played them. David Bamman, Brendan O'Connor and Noah A. Smith, "Learning Latent Personas of Film Characters," In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics 1: Long Papers (ACL: 2013), 352.

[4] i.e. Authors may also employ dialog and description to articulate character traits.

[5] David Bamman, Ted Underwood and Noah A. Smith, "A Bayesian Mixed Effects Model of Literary Character," In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics 1: Long Papers (ACL: 2014), 377.

[6] Bamman et. al., "A Bayesian Mixed Effects Model of Literary Character," 377.

[7] Bamman et. al., "A Bayesian Mixed Effects Model of Literary Character," 377.

[8] A bibliography of the works in our corpus is available online at: doi:10.7910/DVN/3UXBOJ. 250 of the British novels were from the Chadwyck Healey Nineteenth-Century Fiction collection. 900 of the American works were from the Chadwyck-Healey Early American Fiction collections. The rest of the collection was made up of works collected from Project Gutenberg, the internet archive, and ad hoc scanning efforts of the Stanford Literary

Historical Background

In order to study the depiction of male and female characters, we focused on character action as a proxy for what we will call "character agency" or simply "agency." Though there are several elements that inform a character's portrayal, including the way they speak (in dialog) and the author's physical description of the character, examining character action as expressed through verbs offered a practical window into the relationships among gender, characterization, and writerly convention. Since the texts in our corpus were published between 1800 and 1900, we expected to observe trends that correspond with scholarly research on 19th century Western gender conventions. For example, Barbara Welter observes that "submission" was considered a feminine virtue during the mid 19th century. According to Welter, men "were supposed to be religious, although they rarely had time for it, and supposed to be pure, although it came awfully hard to them, but men were the movers, the doers, the actors. Women were the *passive, submissive responders.*"[9] In *The Crisis of Action in Nineteenth-century English Literature,* Stefanie Markovits articulates the way in which these conventions were explicitly tied to action, referring to the Victorian period as characterized by "women's limited sphere of action." [10] To make this claim more explicit, Markovits cites E. S. Dallas, the *Times* literary critic who in 1866 wrote that the "first object of the novelist is to get personages in whom we can be interested; the next is to put them in action. But when women are the chief characters, how are you to set them in motion? The life of women cannot well be described as a life of action." [11] The existence of such conventions led us to believe that in most of the 3,329 novels studied, we would observe male characters behaving differently than female characters, since Victorian notions of propriety stressed the docile, passive, and domestic aspects of feminine behavior.

In their influential study of Victorian gender roles, *The Madwoman in the Attic,* Sandra Gilbert and Susan Gubar note tendencies similar to those observed by Welter. They observe that during the Victorian period women were categorized as either angels or monsters depending on how well they adhered to the valorization of passive, domestic female behavior.[12] Though focused primarily on prescriptive female gender roles, these observations point to the fact that ideal male behavior was also highly standardized. While men may have been categorized as the *doers*, this *doing* was also limited to *appropriate* actions; emotional, domestic, and passive actions would have been considered feminizing. Though the 19th century was characterized by literature that stressed these values, such as etiquette manuals and moralistic novels, the prevalence of female writers during this period highlights the way in which gender roles were also changing. Jane Austen, Charlotte Brontë, George Elliot, Maria Edgeworth, and Ann Radcliffe are just a few of the many female authors writing during this period. Though it would be incorrect to assume that all female writers were working against gender stereotypes, the growing predominance of female authorship, and female readership, points to changing understandings about gendered behavior.[13]

To accomplish our study, we needed a way to track both the appearance of character and the actions associated with characters. We turned toward a previous study, conducted by the University of Nebraska Literary Lab in our efforts to record character action. This study, conducted by Baylog (et. al.), in 2014,[14] examined the connections between gender and action by extracting gendered pronouns and the verbs following them. Gendered pronouns, such as "she" and "he," offer a useful and reliable way of extracting character presence and recording gender. Extracting verbs, such as "to faint" and "to command," when they occurred after a pronoun in a single sentence, provided a straightforward if, at times, imperfect method for exploring the types of behaviors typically associated with characters and for subsequently exploring the ways that 19th century gender stereotypes and expectations were portrayed in literature. Our current research adopted this approach of utilizing gendered pronouns and verbs, but we have expanded and improved on the previous study both by

---

Lab between 2008-2011. The materials from the internet archive and the Stanford scans were uncorrected OCR.

[9] Barbara Welter, "The Cult of True Womanhood: 1820-1860," American Quarterly 18, no. 2 (1966): 151-74. Emphasis added.

[10] Stefanie Markovits, The Crisis of Action in Nineteenth-century English Literature (Columbus: Ohio State University Press, 2006), 103.

[11] Stefanie Markovits, The Crisis of Action in Nineteenth-century English Literature, 103.

[12] Sandra M. Gilbert, and Susan Gubar, The Madwoman in the Attic: The Woman Writer and the Nineteenth-century Literary Imagination (New Haven: Yale University Press, 1979).

[13] For historical information about the advent of female authors and readers, and female writing in general during the 19th century, see: Susan S. Williams, Reclaiming Authorship: Literary Women in America, 1850-1900 (Philadelphia: University of Pennsylvania, 2006); Joanne Shattock, Women and Literature in Britain: 1800-1900 (Cambridge: Cambridge University Press, 2001).

[14] Oliver Baylog, Laura Dimmit, Travis Heller, Gabi Kirilloff, Shannon Smith, Grace Thomas, Chandler Warren, and James Wehrwein, "'More than Custom has Pronounced Necessary': Exploring the Correlation between Gendered Verbs and Character in the 19th Century Novel," UNL Digital Commons, accessed January 22, 2016.

refining the methods for pronoun and verb extraction and by beginning to address how literary genre complicates questions surrounding gender and agency. Among other things, our study explores the extent to which genre may play a role in shaping the sorts of agency that male and female characters are "allowed" to possess.

Questions of Genre

Though useful to our everyday understanding of literature, "genre" is a highly contested concept.[15] While the issues surrounding genre analysis are complicated, we find that a useful strategy for studying genre is offered by an approach that understands these categories as historically and socially constructed. Such an approach inspires questions such as, "what types of themes, tropes, plots, and characters did various authors feel were appropriate (or inappropriate) for the genres they were engaging with?" "When and how did authors depart from genre conventions, or push the boundaries of the genres they were interacting with?" In order to compare overall trends in our corpus of novels with trends in specific genres, we needed a way to determine which genres to examine, and which novels participated in those genres. Initially, we focused on categories, or types of writing, that reflected 19th century, as opposed to contemporary, understandings of genre. We wanted to begin our study by focusing on a relatively distinct and recognizable genre. We turned our attention to the Gothic because there was a 19th century conception of the Gothic novel as a type of writing that was differentiated from the broader and more nebulous categories of "fiction" and "the novel." This is not to say that the term "Gothic" itself was consistently used throughout the course of the 19th century to define this genre. Novels that are now considered to be indicative of the Gothic genre may at the time have been referred to as Romances or supernatural fiction. However, the distinct nature of this category as separate from the nebulous category of fiction at large can be seen in comments from writers, scholars, and reviewers. For example, in 1800 the Marquis De Sades remarks of the Gothic that "this genre was the inevitable product of the revolutionary shock with which the whole of Europe resounded... ."[16] Similarly, in *Northanger Abbey*, a novel that critics have identified as exploring and parodying Gothic conventions, Jane Austen groups seven "horrid books," each of which belongs to the Gothic genre. [17] In addition, the fact that the Gothic genre has been the subject of feminist criticism and debates about gender roles made it especially applicable to our research.[18] In order to determine which books in our corpus are associated with the Gothic genre, we relied on encyclopedias of Gothic literature and scholarly sources to select 33 texts from our corpus that are recognized as strongly indicative of the Gothic genre (See Appendix A). We hypothesized

---

[15]The problems associated with genre are irrevocably tied to the concept's benefits; as a classification scheme, it helps us to talk, think, and write about literature, but it can subsequently impose artificial and socially constructed boundaries on a text. As Adena Rosmarin notes, the conventional uses of genre theory tend to "naturalize or historicize the genre by retrospectively 'finding' it in the literary text…" thus imposing an artificial category on a more fluid, malleable, and indeterminate textual form (Adena Rosmarin, The Power of Genre (Minneapolis: University of Minnesota Press, 1985), 26). In addition, scholars have observed the extent to which single works participate in multiple genres, making acts of categorization difficult. Further complicating these issues is the question of which textual features contribute to our understanding of genre. As Amy Dewitt observes, genre has been traditionally explored in terms of form, rather than content, even though such a divide does not take into account the way in which multiple traits inform our understanding of genre (Amy J. Devitt, "Generalizing about Genre: New Conceptions of an Old Concept," College Composition and Communication 44.4 (1993): 573). Computational analysis offers a useful tool within this conversation precisely because it allows scholars to look for commonalities among a large corpus of texts. Such methodologies allow scholars to search for patterns across a diverse corpus, and to examine whether or not these patterns relate to perceived genre categories. The connection between gender and genre offers rich possibilities for research, since it interrogates the ways in which notions of social propriety influence writing. In speaking of the connection between gender and genre, Marjorie Stone notes, "the Victorian tendency to insist upon the hierarchy of genders may well have contributed to the stubborn persistence of the traditional hierarchy of genres" (Marjorie Stone, "Genre Subversion and Gender Inversion: 'The Princess' and 'Aurora Leigh'," Victorian Poetry 25, no. 2 (1987): 102. Tied to this observation is the understanding that 19th century notions of propriety dictated that certain types of writing and certain topics were considered inappropriate for female writers. In addition, scholars have observed the ways in which specific genres, such as the Gothic novel, demonstrate particular trends in their portrayal of gender. As Donna Heiland notes, "the transgressive acts at the heart of Gothic fiction generally focus on corruption in, or resistance to, the patriarchal structures that shaped the country's political life and its family life, and gender roles within those structures come in for particular scrutiny" (Donna Heiland, Gothic & Gender: An Introduction (Malden: Blackwell, 2004), 5). If certain genres were considered more appropriate for male or female authors, and if different types of gender characterization are associated with specific genres, we might ask if, and how, this manifests itself in the behavior of male and female characters.

[16]De Sade, Idee sur les romans (Paris, 1800), quoted in E.J. Clery, The Rise of Supernatural Fiction (Cambridge: Cambridge University Press, 1999), 156.

[17]Jane Austen, Northanger Abbey (London: CRW Pub., 2004), 44. For more, please see the introductions of Montague Summers, The Gothic Quest; a History of the Gothic Novel (New York: Russell & Russell, 1964).

[18] For an examination of gender in the Gothic novel, see: C. Wynne's Bram Stoker, Dracula and the Victorian Gothic Stage (London: Palgrave Macmillan, 2013), which claims that "Gothic performance opens up a space to challenge sexual and gender norms" (11). The fact that the Gothic novel can be viewed as a space in which to explore Victorian gender roles does not mean that Gothic writers were themselves intentionally challenging these roles. As Jarlath Killeen observes in History of the Gothic: Gothic Literature 1825-1914 (Cardiff: University of Wales Press, 2009), Gothic writers themselves may have been "conservative" even while their fiction "forced them into unearthing some of the limitations of the gender roles ascribed to women in the Victorian period."

that it would be fruitful to explore the possibility of a correlation between the continuities observed by scholars in their categorization of these works and the patterns in characterization that we extracted.

In addition to examining the Gothic genre, we explored the thirty-six 19th century novels that were examined in "Quantitative Formalism"[19] and used in previous experiments to test the ability of computational tools to sort texts based on genre. The corpus from Allison et. al. included novels from the following genres: Gothic (6), historical tale (4), national tale (4), industrial (6), silver-fork (4), Bildungsroman (6), Evangelical (2), Newgate (2), and anti-Jacobin (2). Like our set of 33 Gothic novels, these novels were selected because they are highly indicative of their associated genres. While some of these genres may reflect contemporary, or specifically scholarly conceptions of genre, examining these works allowed us to ask useful questions about the relationships among gender, action, and genre. For example, while authors whose works appear in our selection of industrial novels may not have conceived of their works as specifically "industrial," these categorizations still reflect a scholarly understanding of continuities that exist among these texts. We hypothesized that it would be fruitful to explore if there were a correlation between the continuities observed by scholars in their categorization of these works and the patterns in characterization that we extracted.

Methodology

For this study we defined two classes of pronouns: the male pronouns were "he" and "him," and the female pronouns were "she" and "her." Along with these pronouns, we also included the gendered nouns "man", "woman", "men" and "women." Though not strictly pronouns, the later nouns are clearly gendered and were included in order to expand the number of potential male and female characters being tracked. In what follows, therefore, we use the term "pronoun" loosely in that we include these later nouns along with pronouns such as "he" and "she."

To grammatically parse the sentences in our corpus of novels and thereby identify pronoun verb pairings, we relied on the Stanford dependency parser that is included in the open source Stanford CoreNLP toolkit. [20] The dependency parser identifies which words in a sentence are the subject or object of a verb. Consider the following sentence:

"She watched as her dog ate the rabbit."

The Stanford tool parses the example sentence and correctly identifies that the verb "watched" is connected to the pronoun "She":

nsubj(watched-2, She-1)
root(ROOT-0, watched-2)
mark(ate-6, as-3)
nmod:poss(dog-5, her-4)
nsubj(ate-6, dog-5)
advcl(watched-2, ate-6)
det(rabbit-8, the-7)
dobj(ate-6, rabbit-8)

Below is another example showing how the parser identifies verbs in the gerund form and pairs them with the subject. Here the parser correctly identifies that the verb "listened" is associated with the female pronoun "she" but also that the gerund "talking" is associated with the pronoun "him."

"She listened to him talking."

nsubj(listened-2, She-1)
root(ROOT-0, listened-2)

---

[19]Sarah Allison, Ryan Heuser, Matthew L. Jockers, Franco Moretti, and Michael Witmore, "Quantitative Formalism: An Experiment," N+1, no. 13, (Winter 2012), 81-108.

[20]Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. "Generating Typed Dependency Parses from Phrase Structure Parses," 5th International Conference on Language Resources and Evaluation (2006); Marneffe, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D. Manning, "Universal Stanford dependencies: A cross-linguistic typology," 9th International Conference on Language Resources and Evaluation (2014); Danqi Chen and Christopher D. Manning, "A Fast and Accurate Dependency Parser using Neural Networks," In Proceedings of EMNLP (2014). Also see http://nlp.stanford.edu/software/lex-parser.shtml

```
mark(talking-5, to-3)
nsubj(talking-5, him-4)
advcl(listened-2, talking-5)
```

Dependency parsing every single sentence in the 3,329 novels in our corpus was computationally expensive. Indeed, in similar work, Bamman, Underwood, and Smith (2014) specifically avoided dependency parsing their corpus of 15,000 texts because the approach was "too slow for the scale of [their] data." [21] For this reason, we leveraged the resources of the Holland Compute cluster at the University of Nebraska, where the entire parsing task was completed in less than 24 hours.[22]

Once the dependency parsing was complete, we post-processed the output using basic regular expression scripting to identify and extract the *nsubj* / *verb* pairings that met our pronoun class criteria. In the first example sentence above, the pronoun "she" is identified as the subject of the sentence along with its corresponding verb "watched." Together, these two words form the *nsubj* grouping.

The output of our post processing script was a series of CSV files, one for each input novel. Each CSV file contained four columns indicating the unique text from which the data was extracted ("file_id"), the identified verb ("verb"), the raw count of how many times the verb was associated with a male pronoun ("male_count"), and the raw count of how many times the verb was associated with a female pronoun ("female_count"). An example of the output for the verbs "admired" and "ran" in Grace Aguilar's 1847 novel *Home influence: a Tale for Mothers and Daughters* is shown in Table 1:

file_id	verb	male_count	female_count
2	admired	1	2
2	ran	3	5

Table 1.

In this case there was one instance of a male pronoun paired with the verb "admired" and two instances of a female pronoun paired with the verb "admired." There were three occurrences of a male pronoun paired with the verb "ran" and five instances of a female pronoun paired with "ran."[23]

The data from all of the CSV files was combined into a single long form matrix of dimension 4,314,192 x 4. This long form matrix was reshaped into a wide form matrix of 6,658 rows by 124,190 columns. For each row, an *ID* column indicated the source novel, a *class* column indicated the pronoun gender, and a column for each of 124,190 *verbs* associated with the novel and pronoun gender metadata recorded in the previous columns. There are, therefore, two rows for each novel: one row contains the data relating to female pronouns and the other contains the male pronoun data.[24] The cells in the matrix contain the raw counts of each verb in each novel according to associated pronoun class row.

Pronoun Gender Normalization

In our corpus, male pronouns occur 1.6 times more often than female pronouns. Figure 1 offers a chronological representation of this difference over the course of time represented in our corpus. To deal with this imbalance, it was important to convert all of the raw counts into percentages so that we could understand and compare differences in usage based on pronoun gender unbiased by the fact that there are many more male pronoun pairings overall. To achieve this end, we divided the raw counts of each verb in each novel by the sum of all of the raw counts in each novel so as to create a percentage

---

[21] David Bamman, Ted Underwood and Noah A. Smith, "A Bayesian Mixed Effects Model of Literary Character," 371

[22] Some of the files in our collection contained uncorrected OCR which presented an especially challenging situation for the parser. One text, for example, contained an unpunctuated list of 225 words that the parser believed to be a very long sentence. When we first ran the parser without setting a maximum sentence length flag, the parser repeatedly failed when trying to parse this long list as a sentence. Setting the maximum to 50 resolved this and similar problems generated by poor OCR.

[23] When we speak of "pairs" here we mean the subject of the sentence and the verb that the subject is performing. Thus, in the sentence "She admired him" the subject we are extracting is "she" with the verb "admired."

[24] Obviously, there is nothing special about our decision to store the male and female data in the same matrix. This was done for convenience. The data could have been stored in two separate matrices, one for male pronouns and one for females.

usage of each verb for each pronoun gender. Consider the case of the verb "said" in William Ainsworth's *Merry England Or Nobles And Serfs*. In this text, the verb "said" is associated with a male pronoun 57 times and with a female pronoun 24 times. However, in this text, there are 1319 male pronoun-verb pairings and only 539 female pronoun-verb pairings. If we simply analyzed the raw counts of the verb "said," it might appear that "said" is far more likely to be associated with male pronouns. After we converted the raw values to percentages, however, we found that association of the verb "said" with female pronouns is actually slightly higher as a percentage of all female pronoun verb pairs (4.5% of female pronoun-verb pairings) than it is for males (4.3% of male pronoun-verb pairings). By converting all of the raw counts to frequencies that are relative to the total number of either male or female pronouns in each text, we compensate for the imbalance between occurrences of male and female pronouns.
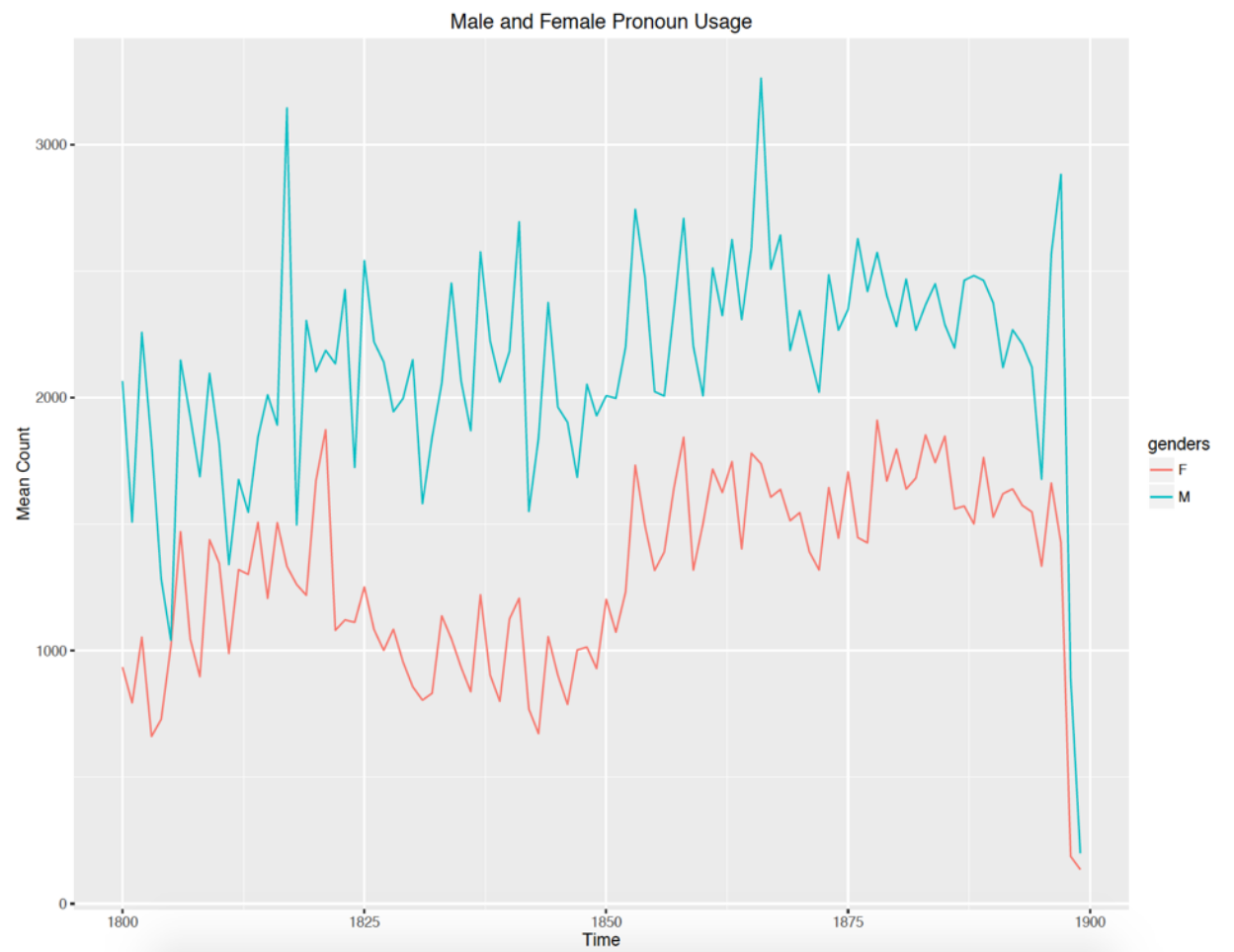

Figure 1. Mean Male and Female Pronoun Occurrence.

Winnowing

The dependency parser was not 100% accurate in its identification of verbs. Human analysis of the parsed data revealed that many words identified as verbs were not in fact verbs. The errors were of two primary types. In the first case, the parser was simply wrong. These mistakes often involved classifying an adjective as a verb. In the second case, the parser identified as a verb a character string that was the result of there being some files in our corpus composed of uncorrected OCR. In these later cases, the parser sometimes identified a "non-word" string of characters that was functioning in the

sentence as a verb, but was not spelled correctly. Since these errors were not systematic across the entire corpus, and because they tended to be infrequent compared to verbs that were correctly identified, we opted to winnow our data to include only those verbs that occurred with a very high frequency across the corpus. This winnowing also gave us the added advantage of limiting our analysis to those verbs that readers are most likely to encounter and that tend to be used in association with either pronoun gender, albeit at different rates of frequency. [25]

Our feature winnowing strategy involved first calculating the mean frequency of each verb across the entire corpus. We then retained only those verbs in the top 0.0025 (1/400). This resulted in a subset of 310 highly frequent verbs. While a study of less frequent verbs is also potentially interesting, we believe that a focus on high-frequency verbs is of particular value in terms of helping us to understand the verb-pronoun patterns that are most consistent and most "established" in the conscious or unconscious minds of the authors who wrote these books.

Even after winnowing in the manner described above, human analysis of the resulting verbs revealed thirty-nine that were either not verbs or were ambiguous enough to warrant exclusion. These thirty-nine tokens (Appendix B) were added to a blacklist and removed from the analysis. The final list of 281 verbs is found in Appendix C.

The normalized and winnowed data was merged with eight columns of book level metadata to create a final matrix of 6,658 x 289.[26] The metadata we catalogued included the book's file name and unique ID, the author's first and last names, the author's biological gender (where known), the book's title, year of publication, and the pronoun class (M/F) corresponding to the verb data in the 281 columns. [27]

Experiments and Observations

Once the data were pre-processed in the manner described, we ran a series of classification experiments using the nearest shrunken centroids classifier.[28] The NSC classifier has the advantage of being a highly interpretable classifier; it not only returns class predictions and the probabilities associated with those predictions, but it performs feature selection and provides statistical data about which features were found to be most useful in the overall classification and which classes those features are most or least associated with.[29] Since our objective was not ultimately to predict likely pronoun genders from a set of verbs, but rather to understand the types of agency associated with male and female characters in the 19th-century novel, NSC was an effective model to deploy since it would distinguish between which verbs showed little value in separating the two gender classes and which verbs showed a strong association with one class or the other.

We ran a series of classification experiments on our data in order to evaluate the strength of the verb to pronoun association and to explore whether there were external factors such as author gender or novel genre that influenced the classifier's ability to predict the gender of a pronoun based on the verbs with which it is most often associated.

Classification

We began with a straightforward 10-fold cross validation experiment using a randomly selected two-thirds of the data. [30] In this experiment the classifier reported an overall accuracy of 81%. When the true pronoun class was female, the machine

---

[25] Another type of study might elect to focus on infrequent verbs or verbs that are only used in association with one pronoun gender or the other.

[26] The normalized and winnowed data is included in the file titled "percentage_data.csv"

[27] We understand that gender can be fluid. Authors were coded as being either male or female based on biographical research. Authors working under known pseudonyms were coded according to their known biological gender and not the implied gender of their pseudonyms. For example, works authored by Mary Ann Evans under the pseudonym George Elliot were coded as female. In cases where a biological gender could not be determined, author genders were marked as "unknown."

[28] Robert Tibshirani, Trevor Hastie, Balasubramanian Narasimhan, and Gilbert Chu, "Diagnosis of multiple cancer types by shrunken centroids of gene expression," PNAS 99, no. 10 (2002): 6567-6572; 2003; Tibshirani, Trevor Hastie, Balasubramanian Narasimhan, and Gilbert Chu, "Class Prediction by Nearest Shrunken Centroids, with Applications to DNA Microarrays," Statistical Science 18, no. 1 (2003): 104-117; Matthew L. Jockers and Daniela M. Witten, "A Comparative Study of Machine Learning Methods for Authorship Attribution," Literary and Linguistic Computing 25.2 (2010): 215-224.

[29] Appendix D provides a list of the top 100 male and female oriented verbs that the machine identified as being important during one random run of the classifier.

[30] We performed a series of runs using different random selections, but for the sake of repeatability, the data reported here were obtained using a set random seed of 1966. All of the code for this classification is found in the file titled "classification.R"

reported a 22% error rate in precision and when the true pronoun class was male, a precision error rate of 16%. The rate of error observed in cross-validation of the training data was closely reflected in the prediction accuracy observed when the model was tested on 658 randomly selected rows that were held out prior to training. In the test involving held out, unseen, data, 18% of male pronouns were mistakenly classified as female and 23% of female pronouns were misclassified as male, an overall error rate of 21%.[31]

This 30% improvement over chance suggested that there was indeed a strong association of certain verbs with pronoun gender. In one round of random sampling, the classifier identified that the verbs "wept," "sat," and "felt" were strongly associated with female pronouns and the verbs "took," "walked," and "rode" were strongly associated with male pronouns. The distribution of these verbs appeared to support existing scholarship on 19th century gender stereotypes; many of the verbs indicating a female pronoun are associated with emotion while many of the verbs indicating a male pronoun are associated with physical action and motion.

After conducting a series of 10-fold cross validations in the manner described above, we also conducted a large-scale, hold-one-out validation in which each text was successively held out, while a new model was trained on all of the remaining data and then used to predict the class of the held out data. Ten verbs (five male and five female) the machine found most useful in differentiating between male and female pronouns—averaged across all of the 3,329 hold one out cross-validation test—are listed in Table 2. The rows are ranked from one to ten based on how useful they were to the model. The third column of the table indicates the gender of the pronoun most typically associated with the verb shown in the second column.

Rank	Verb	Pronoun Association
1	took	M
2	wept	F
3	sat	F
4	looked	F
5	felt	F
6	cried	F
7	found	M
8	rode	M
9	walked	M
10	come	M

Table 2.

In the hold one out experiment, the machine correctly classified pronoun gender 5,389 out of 6,658 times for an overall accuracy of 81%. In terms of accuracy, this result was identical to the result observed in the single 10-fold, randomly sampled cross validation experiments. In the hold one out tests, when the machine guessed the class incorrectly, 1,269 times, it was almost twice as likely to guess wrong when the true class was female. 738 of the 1,269 incorrect classifications were for female pronouns whereas there were only 531 male pronouns that were misclassified as being female. This result may indicate that there is less variation when it comes to verbs associated with male characters, and it suggests that the verbs most indicative of female pronouns may be less stable, or more variable, than those associated with male pronouns. One might go so far as to say that female pronouns are slightly less gendered or perhaps less "codified" whereas the male pronouns seems to occupy a circumscribed behavioral space. What we observe, therefore, is a situation in which a verb typically associated with a female pronoun, such as "acknowledge," is still used in association with some male pronouns, while a verb typically associated with male pronouns, such as "rode," is three times less likely to be used in conjunction with a female pronoun.

In addition to examining the accuracy of the class predictions, NSC provides class probabilities that can be examined to assess the level of confidence the classifier has in its class predictions. An analysis of the probabilities revealed that when the machine guessed the class correctly, on average it was 76% confident when the true pronoun was male and 80% when the true pronoun was female. When the machine guessed incorrectly, it was less confident (33% mean probability) when assigning female as the class than when assigning male (37% mean probability). These average probabilities indicate that the machine is less confident in its assertions about verbs associated with female pronouns. By extension, this suggests

---

[31]The classes were evenly balanced, so we can assume a baseline accuracy of 0.5.

that the verbs associated with female pronouns are generally more ambiguous in terms of projecting a clear pronoun gender class.

The classification results indicated that for the corpus as a whole there were strong associations between verbs and pronoun gender. Our next task was to segment the corpus according to genre and determine whether or not the overall prediction accuracies were sustained. The model achieved 58% accuracy predicting the pronoun gender in our six Bildungsroman novels, 63% accuracy in four silver-fork novels, and 67% accuracy in three historical novels. Accuracy of 75% was observed for our two anti-Jacobin, four Evangelical, and eight national tale novels. Across thirty-three Gothic novels, we observed 80% accuracy and 100% accuracy for our six industrial and two Newgate novels (see figure 2).
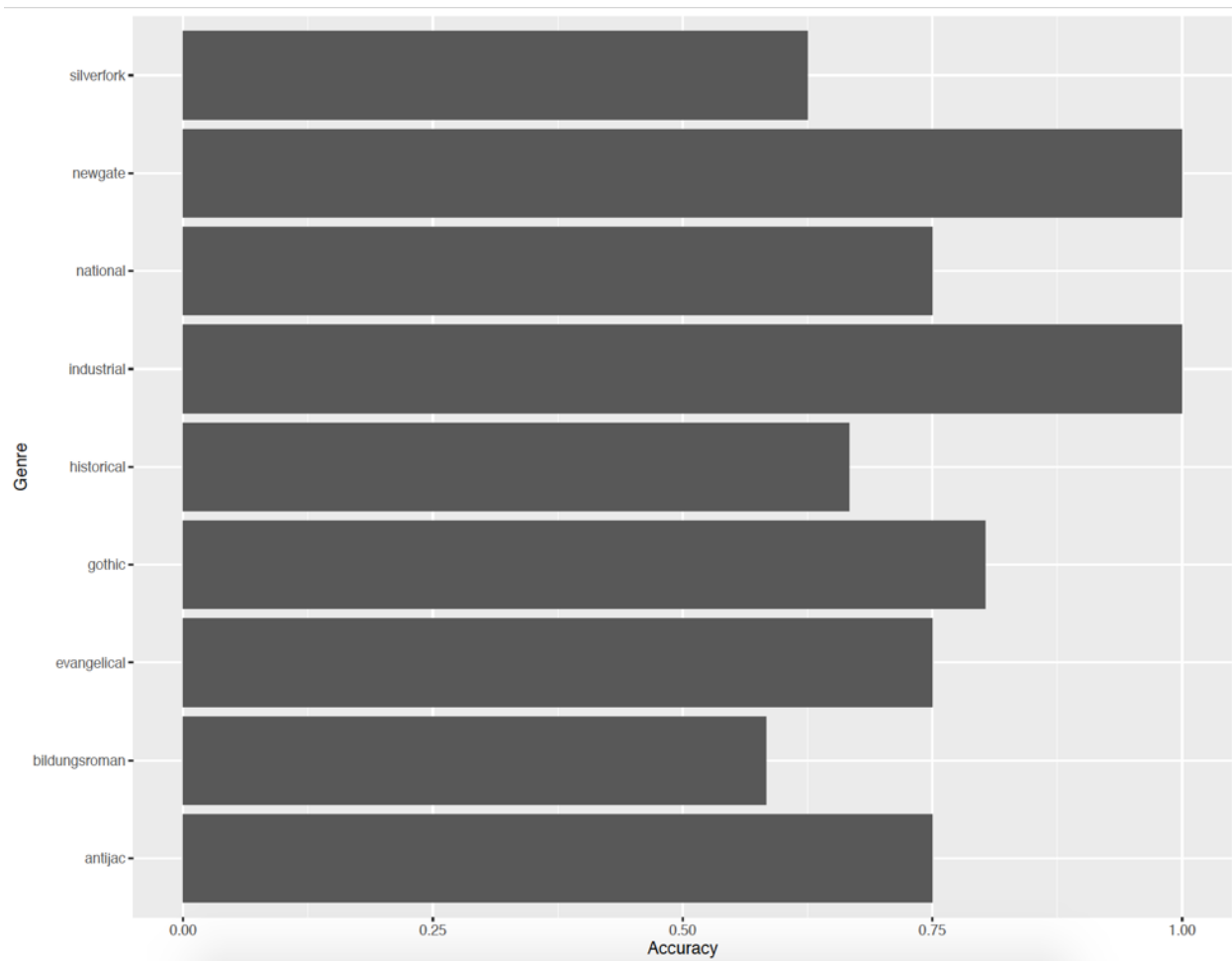

Figure 2.

The small number of novels examined means that these accuracy figures are far from significant. However, it is interesting to consider the possible ways in which Newgate and industrial novels may be more strictly codifying gender than what was seen in the Bildungsroman and historical novel. The evidence, albeit scant, suggests that the male and female characters found in Newgate and Industrial novels tend to be more consistently "cast" and distinctly different from each other in terms of the types of agency they are assigned, whereas in the other two genres more blending of agency is observed (i.e. male and female characters were assigned more flexibility in terms of crossing gender agency norms). Such an interpretation does make some sense when we consider that the criminals in Newgate novels have fairly limited and restricted agency. Something similar might be said of workers in the industrial novel. Whereas a Bildungsroman or historical novel may present characters at different stages of life and, therefore, characters whose agency changes/evolves over time and is not static.

While our lack of extensive genre metadata meant that we could not form robust conclusions about the relationship between character agency and genre, we did have complete metadata pertaining to author gender, so it was natural to explore whether there was a detectable relationship between classification accuracy and author gender. Author gender was not found to be

a strong determiner of classification accuracy. In our metadata, we have three classes of author gender: male, female, and unknown. The results observed within each of these classes mirrored what we observed across the entire corpus. Table 3 provides a summary of the data when broken down by author gender.

Author Gender	Incorrect Predictions	Correct Predictions
F	555 (20%)	2153 (80%)
M	624 (18%)	2868 (82%)
U	90 (20%)	368 (80%)

Table 3.

In the face of this result, we speculated that perhaps there might be a relationship between author gender and one particular pronoun gender or the other. To explore this, we computed the extent to which the machine guessed correctly when the author was male and the pronoun was male vs. when the pronoun was female. We did the same for female authors. The results are reported in Table 4.

Author Gender	Pronoun Class	Incorrect Predictions	Correct Predictions
F	F	258 (19%)	1096 (81%)
F	M	297 (22%)	1057 (78%)
M	F	420 (24%)	1326 (76%)
M	M	204 (12%)	1542 (88%)
U	F	60 (26%)	169 (74%)
U	M	30 (13%)	199 (87%)

Table 4.

The classification data reported in Table 4 suggest that the machine has a more difficult time classifying male characters when the author is female and female characters when the author is male. In addition, female authors were likely to create male characters who acted outside of the overall trends we have observed for the corpus as a whole. Similarly, male authors were more likely to create female characters that defy the trends observed throughout. This may suggest that male and female authors were more conventional when creating characters of their own gender. However, our results also suggest that this trend is more indicative of male authors than female authors. The classification error rate for male characters created by male authors is only 12%, indicating that male authors created male characters that behaved in accordance with the overall trends far more consistently than when writing female characters.

We initially expected to discover the opposite of this result; we had hypothesized that female authors would be more likely to create female characters that challenged gender stereotypes by acting in a more "masculine" fashion. What we observed, however, indicates that the female authors are associating male pronouns with verbs in a way that is not quite in line with the overall use of "male-oriented" verbs in the larger corpus. Additionally, the machine has a more difficult time classifying the male usage of female pronouns which suggests that the male authors are diverging from the large scale patterns of female pronoun to verb usage.[32] Finally, when the author gender is unknown, we observe that the machine struggles a bit more with predictions of female pronouns.

Next we explored the data in order to identify outlier books. There were nine books in which the model guessed wrong for both pronoun classes: These were as follows: Matthew Lewis's *Romantic Tales,* Thomas de Quincey's *Klosterheim Or The Masque,* Charlotte Brontë's *Jane Eyre,* Amelia Edwards's *Miss Carew,* Douglas Jerrold's *Men Of Character,* Julie Chetwynd's *A Brilliant Woman,* Wilkie Collins's *The Law And The Lady*, Allan Cunningham's *Sir Michael Scott A Romance*, and Charles Maturin's *Melmoth The Wanderer A Tale*. We elected to explore Maturin's *Melmoth the Wanderer* and Collins's *The Law And The Lady* at a slightly closer scale because *Melmoth* was the worst scoring novel with a central male protagonist and *The Law And The Lady* was the worst for a novel with a central female protagonist. In Maturin's novel, a large number of verbs that our model identified as being more typically associated with female pronouns, including

---

[32]One can't help but wonder if these differences are the result of authors not fully understanding the gender conventions associated with the gender opposite to their own.

verbs such as *cried, exclaimed, felt, heard, listened, looked, loved, trembled, wept*, and *whispered*, are disproportionately associated with male pronouns.[33] Here are a few examples in context:

- "Love," he *cried*, extending his arm towards the dim and troubled sky. . .
- He *exclaimed*, "What is the matter? You have alarmed me by your cries,—you pronounced the name of the infernal spirit,—what have you seen? what is it you fear?"
- However it was, he *felt* himself compelled to tell her it was a new religion, the religion of Christ, whose rites and worshippers she beheld.
- —while he *heard* accents issue from those lips which he *felt* it would be as impossible to pervert as it would be to teach the nightingale blasphemy,—he *sunk* down beside her, passed his hand over his livid brow, and, wiping off some cold drops, thought for a moment he was not the Cain of the moral world, and that the brand was effaced,—at least for a moment.
- Melmoth *felt* as if he *listened* to some herald of "fate and fear."
- He *looked* at her as she fluttered round him with outspread arms and dancing eyes; and *sighed*, while she welcomed him in tones of such wild sweetness, as suited a being who had hitherto conversed with nothing but the melody of birds and the murmur of waters.
- while he, without uttering a word, leaned against the pillars of her balcony, or the trunk of the giant myrtle-tree, which cast the shade he *loved*, even by night, over his portentous expression
- At these words he *trembled*.
- He sat down and *wept*.
- He *whispered* in the softest tones.

Also in *Melmoth the Wanderer*, a number of typically male-oriented verbs were often associated with female pronouns. These included:

- She *entered* the room where they were, and they turned towards her with their usual smiling demand for her approbation.
- When she *arrived* at a certain distance from the Castle, she dismissed the family carriage, and said she would go on foot with her female servant to the farmhouse where horses were awaiting her.
- Again she *bowed* in the deep prostration which indicates the united submission of soul and body, and remained in this attitude of absorption so long, that, when she rose, she did not perceive the absence of her companion.
- She *came*, and, on her introduction to Melmoth, it was curious to observe the mingled look of servility and command, the result of the habits of her life, which was alternately one of abject mendicity, and of arrogant but clever imposture.
- It was remembered, or reported, that she had *made* many efforts to soften the heart and open the hand of her brother
- "I will go back to Germany," she repeated; and, rising, she actually *took* three or four firm and equal steps on the floor, while no one attempted to approach her.

Similar deviations from the norm were observed in *The Law and the Lady*. Verbs more typically associated with female pronouns included, among many others:

- "Hang me, innocent as I am!" he *cried*.
- His expression suddenly changed; his face darkened and hardened very strangely. "Stop!" he *cried*, before I could answer him.
- He *answered* without looking at her; his changeless eyes still fixed, as it seemed, on something far away.

---

[33] It is not entirely surprising to find Maturin and Collins identified as outliers given how both authors have received attention from scholars interested in the portrayal of gender. Scholars have noted, for example, that the depiction of female characters in Melmoth can be read as challenging societal norms. In "The Ethics of Excess in Melmoth the Wanderer" Nathaniel Leach notes that Melmoth depicts female characters in a way that differs from traditional romantic narratives. He argues that the character Immalee is depicted as more than a textbook damsel in distress. In "Experimentation and 'horrid Curiosity' in Maturin's Melmoth the Wanderer" Amy Smith also observes that the female characters in Melmoth are allocated more agency than many of the male characters. In discussing Collins, feminist critics have disagreed on Collins's engagement with Victorian gender norms. In "Wilkie Collin's Modern Snow White (Arsenic Consumption and Ghastly Complexions in The Law and the Lady)" Laurence Talairach-Vielmas states that in "The Law and the Lady, Collins continues his critique of woman's commodification…" However, in "Reading Faces: Physiognomy and the Depiction of the Heroine in the Fiction of Wilkie Collins" Jessica Cox concludes that "In spite of the fact that sensation fiction has frequently been identified as subversive, Collins's novels are in fact littered with depictions of the conventional Victorian feminine ideal" (Jessica Cox, "A Brief Explication of "Reading Faces: Physiognomy and the Depiction of the Heroine in the Fiction of Wilkie Collins," In Wilkie Collins: Interdisciplinary Essays, ed. Andrew Mangham (Newcastle: Cambridge Scholars Publishing, 2007), 111). Collin's overall approach to characterization is also important in thinking about the role of gender in The Law and the Lady. According to Bradley Booth "Collins advised his readers that 'the actions of human beings are not invariably governed by the laws of pure reason.' He also cautions us that we must not take our own limited experience as the criterion of the possible" (Bradford A. Booth, "Wilkie Collins and the Art of Fiction," Nineteenth Century Fiction 6.2 (1951): 140-141).

- He *burst* into tears.
- He had driven the woman whom he *loved* to the last dreadful refuge of death by suicide!

Verbs that were more indicative of male pronouns were frequently associated with female pronouns, as in the following:

- With perfect tact and kindness she *entered* into conversation with me.
- Ascending another step on the social ladder, she *took* her stand on the platform of patronage, and charitably looked down on me as an object of pity.
- When she *ordered* her husband and witness to leave the room, on the day of her death.
- She *left* the carriage in the road; and got into the house by way of the garden — without being discovered, this time, by Dexter, or by anybody.
- She is away at some wonderful baths in Hungary, or Bohemia (I don't remember which) — and where she will go, or what she will *do*, next, it Is perfectly impossible to say.

We also queried the data to identify books that most defied the corpus norms: which books according to these macroscale patterns of verb usage had the most "masculine" female characters and which had the most "feminine" male characters. Henry Hebert's novel *The Warwick Woodlands Or Things As They Were There Ten Years Ago* was identified as the male-authored novel with the most masculine behaving female characters. George Lippard's novel, *Adrian The Neophyte* was identified as having the most feminine behaving male characters. Mary Young's *Right And Wrong Or The Kinsmen Of Naples A Romantic Story* showed female pronouns with associations to verbs strongly indicative of the male pronoun class. Dinah Craik's *Romantic Tales* was identified as having male pronouns that were strongly associated with verbs more typical of the female pronoun class.

Two, more familiar, novels that exhibited similar behavior were *The Picture of Dorian Grey* by Oscar Wilde and *Jane Eyre* by Charlotte Brontë. The male pronouns in *The Picture of Dorian Grey* are frequently misclassified as being female and the female pronouns in *Jane Eyre* are often misclassified as male. When we examine the specific instances, we see, for example, that male actors in *Dorian Grey* are frequently paired with the verb "cried" (18 instances compared to 11 with female pronouns), and "felt" (15 male instances compared with four female invocations), as well as a number of other verbs typically associated with female pronouns: *answered, exclaimed, heard, looked, murmured*, and *sat*. The opposite is seen in *Jane Eyre* where female actors are very often seen *calling*, *leaving*, *proceeding*, and *taking*, all activities more typically associated with male pronouns in our corpus.

Conclusions

By the measures employed in this research, we have learned that there is a strong correlation between character gender and verbs in the 19th century novel. This finding indicates that representations of behavior, or agency, understood in terms of the kinds of actions that are associated with particular pronouns, are an important element of characterization. While other elements of character identity, such as speech patterns and visual appearance, are also important aspects of characterization, our study indicates that what characters are *doing* is a key component of how we understand them. The limited nature of our metadata pertaining to genre, meant that we could not determine if genre is or is not a strong determining factor in the relationship between pronouns and verbs. However, we did observe some anecdotal correlations in a few specific genres that seem to make sense when considered in terms of the kind of character agency we might reasonably expect for these particular genres. Because of the difficulties associated with building a large and robust sample of novels from different genres, it is also possible that our results would change given a larger sample set.

Since our study observed a strong correlation between pronoun gender and verbs, we can conclude that there is evidence for an overarching trend of ascribing certain actions to certain genders in the 19th century novel. Further, it would appear as if this trend corresponds with general notions of gender propriety; we observed that verbs connoting emotion and sentiment (such as to cry, to love, to weep, etc.) were more strongly associated with female characters while verbs connoting action and motion were more strongly associated with male characters (to advance, to approach, to ride, etc.) This result would seem to support the work of literary and cultural critics who have observed the 19th century tendency to valorize passive women and active men. [34] Nevertheless, "passive" is clearly not the best descriptor for many verbs associated with female

---

[34]Using verbs to study the relationship between agency and gender is not unprecedented; in a 1997 study on the psychological perception of agency, Marianne Lafrance, Hiram Brownell and Eugene Hahn observe that different types of verbs imply different degrees of power. This study is particularly

pronouns. While verbs such as "smiled" and "loved" do not denote aggressive physical action, a good argument can be made that they are not describing passive actions either.[35]

In addition, it is important to note that not every pronoun/verb pair that corresponds with the trends we observed is an example of a gendered stereotype. There are three key factors of gendered characterization that our study did not take into account, the first of which is irony. The sincere presentation of a female heroine who is emotional, domestic, and passive is quite different from the intentional parody of such a stereotype. However, in either case, the female character may be associated with the same actions. Jane Austen offers a compelling and easily recognizable example of this phenomenon. In our study, the relationship between gendered pronouns and verbs in *Pride and Prejudice, Mansfield Park, Emma, Sense and Sensibility, Northanger Abbey,* and *Persuasion* corresponded with the overall trends we observed in our entire corpus. This is not to say, however, that these books necessarily subscribe to conventional notions of gender since, as recent critics have argued, Austen's subversive potential lies under a shallow veneer of convention. As Marvin Mudrick observes in *Jane Austen: Irony as Defense and Discovery,* Austen used irony to "expose the incongruities between form and fact, all the delusions intrinsic to conventional art and conventional society"[36]. Thus within the scope of our study, novels in which the author is intentionally pushing against, playing with, or directly critiquing gendered stereotypes may be computationally indistinguishable from works that are perpetuating the same stereotypes.

The second way in which our model may overlook certain aspects of gendered characterization is by not accounting for all forms of character presence. Our work only examines gendered pronouns and not, for example, first person pronouns and proper names.[37] In addition, by looking at male and female pronouns, our study groups characters together, regardless of their status, importance, or emotional valence. For example, when we note that female characters are often associated with emotional verbs, we are speaking of *all* female characters: princesses, witches, maidens, mothers, queens, beggars, etc. Similarly, when we examine a novel in which the characterization of male pronouns does *not* follow the general trends of our corpus, there is no way for us to determine whether all of the male characters are behaving in an unconventional way, or whether it is simply the protagonist, or antagonist. These distinctions are ultimately quite important for a study of characterization and gender, since the implications of creating a laudable female heroine who behaves in a "masculine" manner are quite different from the implications of creating a female villain whose "male" behavior only serves to further demonize her. This last issue may also point towards a potential reason why our investigation of specific genres, including the Gothic novel, did not produce results that differed greatly from the overall trends in our corpus. It is quite possible that *specific* character archetypes, such as the heroine, do behave differently in certain genres, but that in averaging character behavior, this information is lost.

A third item that our method did not account for is narrative time. It is entirely possible that, in a Bildungsroman for example, a character matures over the course of the narrative and that the verbs associated with that character change in accordance with that maturation. Tracking the presence, or not, of such verbal change seems like an interesting avenue of future investigation. Of course, not all narratives move from beginning to end in chronological time, so such investigation is by no means straightforward.

Further exploring the ages and types of characters that are associated with specific behaviors would enrich our study of gender and agency and would allow us to examine other factors that may be at play in creating stereotypes. Rather than assume that the relationship between gendered pronouns and verbs that we have observed is evidence only of gender stereotypes, we might ask how other factors, such as character age, race, class, and profession are tied to character behavior. It is also worthwhile to note that even within the scope of our current work, the correlation between gendered pronouns and verbs is

---

useful to our enterprise because of the way that it is focused on the relationship between text and perception. Hiram, Brownell, and Hahn comment on the unique way that grammar and word choice can impact a reader's, or listener's, perception of power relations. The authors note that "verb type and gender stereotype combine to affect people's perception about who is perceived to bring about interpersonal events..." (1). The authors argue that when action verbs are used, such as "to walk" or "to ride," it is assumed that the subject of the sentence causes the action. However, when a verb describing an emotional state is used, such as "to love" it is assumed that the object of the sentence somehow elicited the emotion. Thus, the subject of an action verb, and the recipient of an emotion verb, are both perceived as causal and assumed to have a high degree of agency. Even though this study was conducted recently, when combined with notions of Victorian gender stereotypes, Hiram, Brownell, and Hahn's study points towards an interesting line of reasoning: if the ideal Victorian woman was associated with acts of emotion, such as loving, feeling, and worrying, than even though 19th century female characters may be associated with a large number of such verbs, these verbs do not necessarily imply the same degree of agency as verbs related to physical action. See Baylog, et. al., "'More than Custom has Pronounced Necessary'," 4-5.

[35]In reviewing a draft of this paper, Ted Underwood noted how a word such as "cried" embodies this precise ambiguity in that it can mean either "weep" or "exclaim" depending upon context.

[36]Marvin Mudrick, Jane Austen: Irony as Defense and Discovery (Berkeley: University of California, 1968), 1.

[37]A first person narrative, employing the pronouns "I" and "me" could be either male or female and might be associated with a different set of verbs than those we identified as being strongly associated with male or female pronouns.

likely indicative of not just gender stereotypes, but also literary stereotypes and archetypes. Specific character types, such as the Byronic hero or the damsel in distress are a reflection of multiple markers of identity, not simply gender.

Acknowledgments

Appendix A:

List of Genre Novels

Last	First	Title
Moore	John	Mordaunt Sketches Of Life Characters And Manners In Various Countries Including The Memoirs Of A French I
Opie	Amelia	Adeline Mowbray Or The Mother And Daughter
Bronte	Charlotte	Jane Eyre
Dickens	Charles	David Copperfield
Dickens	Charles	Great Expectations
Evans	Marian	Daniel Deronda
Evans	Marian	Middlemarch A Study Of Provincial Life
Thackeray	William	The History Of Pendennis His Fortunes And Misfortunes His Friends And His Greatest Enemy
Brunton	Mary	Selfcontrol A Novel Second Edition
More	Hannah	Coelebs In Search Of A Wife Comprehending Observations On Domestic Habits And Manners Religion And Mo
Brown	Charles	Wieland
Byrne	Charlotte	Zofloya Or The Moor A Romance Of The Fifteenth Century
Collins	Wilkie	The Woman In White
Dickens	Charles	The Mystery Of Edwin Drood
Golland	C D	The Ruins Of Ruthvale Abbey A Novel
Hogg	James	The Private Memoirs And Confessions Of A Justified Sinner
Le Fanu	Joseph	Carmilla
Le Fanu	Joseph	In A Glass Darkly
Le Fanu	Joseph	Madam Crowls Ghost And The Dead Sexton
Le Fanu	Joseph	The Evil Guest
Le Fanu	Joseph	Uncle Silas
Le Fanu	Joseph	Uncle Silas A Tale Of Bartram Haugh
Lewis	Matthew	The Monk A Romance
Maturin	Charles	Melmoth The Wanderer A Tale
Mitchell	Isaac	The Asylum Or Alonzo And Melissa An American Tale Foundedon Fact
Moore	George	Grasville Abbey A Romance
Moore	John	Zeluco Various Views Of Human Nature Taken From Life And Manners Foreign And Domestic
Peacock	Thomas	Nightmare Abbey
Poe	Edgar	Tales Of The Grotesque And Arabesque
Polidori	John	The Vampyre A Tale
Quintin	Poynet	The Wizard Priest And The Witch
Radcliffe	Ann	A Sicilian Romance
Radcliffe	Ann	The Italian Or The Confessional Of The Black Penitents A Romance
Radcliffe	Ann	The Mysteries Of Udolpho A Romance Interspersed With Some Pieces Of Poetry
Radcliffe	Ann	The Romance Of The Forest Interspersed With Some Pieces Of Poetry
Reeve	Clara	The Old English Baron

Last	First	Title
Shelley	Mary	Frankenstein Or The Modern Prometheus
Shelley	Percy	Zastrozzi A Romance
Smith	Charlotte	The Old Manor House A Novel
Stevenson	Robert	Strange Case Of Dr Jekyll And Mr Hyde
Stoker	Abraham	Dracula
Thomas	Henry	The Irish Necromancer
Wilde	Oscar	The Picture Of Dorian Gray
Galt	John	The Entail Or The Lairds Of Grippy
Scott	Walter	Waverley
Scott	Walter	Ivanhoe A Romance In The Waverley Novels
Shelley	Mary	Valperga Or The Life And Adventures Of Castruccio Prince Of Lucca
Bronte	Charlotte	Shirley A Tale
Dickens	Charles	Hard Times For These Times
Gaskell	Elizabeth	Mary Barton A Tale Of Manchester Life
Gaskell	Elizabeth	North And South
Kingsley	Charles	Alton Locke Tailor And Poet An Autobiography
Trollope	Frances	The Life And Adventures Of Michael Armstrong The Factory Boy
Edgeworth	Maria	Castle Rackrent An Hibernian Tale Taken From Facts And From The Manners Of The Irish Squires Before The Y
Edgeworth	Maria	The Absentee In Tales Of Fashionable Life
Ferrier	Susan	Marriage A Novel
Morgan	Sydney	The Wild Irish Girl A National Tale
Ainsworth	William	Jack Sheppard A Romance
Lytton	Edward	Eugene Aram A Tale
Disraeli	Benjamin	Vivian Grey A New Edition
Gore	Catherine	Mrs Armytage Or Female Domination
Lamb	Caroline	Glenarvon
Lytton	Edward	Pelham Or The Adventures Of A Gentleman

Appendix B:

Blacklist of words: able, about, afraid, alone, aware, beautiful, better, child, con, creature, daughter, dead, done, fellow, friend, girl, glad, good, guilty, happy, herself, ill, kind, lady, man, more, much, one, pale, person, re, right, silent, w, wife, woman, X..19, X.s, young

Appendix C:

accept, added, addressed, advanced, allowed, answer, answered, anxious, appear, appeared, approached, arrived, ask, asked, be, bear, became, become, been, began, beheld, believe, believed, bent, bore, bowed, bring, broke, brought, burst, call, called, came, care, cared, carried, cast, caught, ceased, chose, closed, come, comes, coming, concluded, considered, continued, could, cried, dared, declared, desired, determined, did, die, died, discovered, do, does, doing, drew, dropped, endeavoured, enter, entered, exclaimed, expected, expressed, face, fallen, fancied, feared, feel, fell, felt, find, finished, follow, followed, forget, forgot, forgotten, found, free, gained, gave, gazed, get, give, given, glanced, go, goes, going, gone, got, grew, had, has, hath, have, hear, heard, held, help, hesitated, hold, hoped, hurried, imagined, inquired, intended, is, keep, kept, kissed, knew, know, known, knows, laid, laughed, lay, leaned, learned, leave, led, left, let, lifted, like, liked, likes, listen, listened, live, lived, lives, look, looked, looking, looks, lost, love, loved, loves, made, make, makes, married, marry, meant, meet, met, moved, murmured, muttered, observed, offered, opened, ought, paid, pass, passed, paused, pay, perceived, placed, played, pleased, pointed, possessed, pressed, proceeded, promised, put, quitted, raised, ran, reached, read, ready, receive, received, recovered, refused, regarded, remain, remained, remembered, repeated, replied, resolved, resumed, retired, return, returned, rode, rose, run, rushed, said, sank, sat, saw, say, says, see, seem, seemed, seems, seen,

sees, seized, send, sent, set, shook, show, showed, sighed, sit, sitting, smiled, sought, speak, spent, spoke, spoken, sprang, stand, started, stay, stood, stopped, succeeded, suffered, supposed, sure, take, taken, takes, talk, talked, tell, think, thinking, thinks, thought, threw, told, took, tried, try, turn, turned, understand, understood, used, uttered, wait, waited, walk, walked, want, wanted, wants, was, watched, went, wept, were, whispered, will, wish, wished, wore, would, write, written, wrote

Appendix D:

List of top 50 verbs typically observed with male and female pronouns (in order of relevance from most relevant to least).

Male: took, found, rode, walked, come, muttered, left, paid, made, get, called, came, do, take, got, find, comes, arrived, call, bowed, taken, intended, set, proceeded, make, offered, meant, led, seized, entered, pay, reached, returned, return, addressed, going, brought, knows, went, give, wrote, succeeded, goes, says, sent, thinks, gained, wants, followed, talked.

Female: sat, wept, looked, felt, heard, cried, loved, knew, burst, answered, rose, love, murmured, listened, sank, ran, exclaimed, smiled, replied, whispered, accept, face, refused, sighed, hear, sitting, was, feared, moved, marry, recovered, bear, kissed, suffered, listen, read, loves, look, threw, lay, sit, uttered, remain, know, feel, tried, receive, understand, said, looks.